矿产预测的统计决策方法

殷 棣 贺国信

本文介绍了矿产预测的一种方法——统计决策分析。文中给出了矿产预测条件下统计决策问题的陈述，该问题在先验概率已知和未知时的两种解法：贝叶斯解和极小极大解(对于 $m > 2$ 和 $m = 2$ 两种情况)，原始数据预处理的方法。文章通过在三〇七五地区的应用实例展示了该方法的应用效果。

矿产预测的基本问题是，在矿产无法直接找到的情况下，地质人员根据对成矿地质环境(控矿条件)的研究，做出研究区成矿远景类别的判断。从数学观点来看，这是一个统计决策问题。这里，研究区的成矿远景类别视为大自然的状态，地质人员对远景的预测视为决策人的决策。由此可见，矿产预测问题可以用统计决策方法来解决。统计决策问题是阵地博奕的一种特定情况。关于阵地博奕、统计决策方法的理论和它们在地质学中的应用有许多专著论述[1-4]。本文拟扼要地介绍一下离散型统计决策方法，并给出它在矿产预测中的应用实例。

一、方 法 概 述

1. 统计决策问题的陈述

设大自然有 m 个状态，例如在预测问题中有 m 个远景类别，其号码分别为 1, 2,…, m. 决策人可以取 m 个决策(即地质人员对远景类别的判断)，分别为 1, 2, · · ·, m。

当大自然的真实状态为 i，而决策人取第 j 种决策时，他所蒙受的损失可以下列方阵表示：

$$L = \begin{pmatrix} l_{11} & l_{12} & \cdots\cdots & l_{1j} & \cdots\cdots & l_{1m} \\ l_{21} & l_{22} & \cdots\cdots & l_{2j} & \cdots\cdots & l_{2m} \\ \cdots & \cdots & \cdots\cdots & \cdots & \cdots\cdots & \cdots \\ l_{i1} & l_{i2} & \cdots\cdots & l_{ij} & \cdots\cdots & l_{im} \\ \cdots & \cdots & \cdots\cdots & \cdots & \cdots\cdots & \cdots \\ l_{m1} & l_{m2} & \cdots\cdots & l_{mj} & \cdots\cdots & l_{mm} \end{pmatrix}$$

L 称为损失函数。

由于大自然的真实状态是未知的，又不能直接观测到，决策人为了做出正确的决策，就必须收集一定数量的与大自然状态有关的信息。这就要对与大自然状态有关的 n 种可观测变量进行观测，用以描述大自然的状态。也就是说，地质人员必须事先对与成矿有关的 n 种地质变量进行观测。设每种变量均可取 k 个离散值。那么，每次观测结果均为一个 n 维向量 $\mathbf{x} = (x_1, x_2, \cdots\cdots, x_n)$。所有可能的这种向量共有 $l = k^n$ 个，其集合以 X

本文 1984 年 10 月收到，1985 年 4 月改回，季国容编辑。

表示。

对于大自然的每种状态，上述 n 维向量$\mathbf{x}$均有自己的概率分布 $P_i(\mathbf{x})$，$i=1, 2, \cdots, m$。这样的概率分布族以 $\{P_i\}$ 表示。

大自然的m种状态的概率分布称为先验概率，以 $\mathbf{q}=(q_1, q_2, \cdots, q_m)$ 表示，其集合以Q表示。

对于每个可能的向量 $\mathbf{x}\in X$ 都给出m个决策的某种概率分布 $\varphi(\mathbf{x})=(\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \cdots, \varphi_m(\mathbf{x}))$ 的函数φ称为随机决策函数。它是用以指导决策人根据观测向量 $\mathbf{x}$ 做出某种决策的工具，故又称为决策人的行为策略。其集合以Φ表示。

当大自然取第 i 种状态，决策人选用决策函数φ时，决策人所冒的风险以加权平均损失来表示，即

$$\rho(i, \varphi)=\sum_{j=1}^{m}\int_X L(i, j)\varphi_j(\mathbf{x})dP_i(\mathbf{x}) \tag{1}$$

ρ称为风险函数。

现在可以给出统计决策问题的一种被称为“混合扩充”的陈述了：给定条件组 (Q, Φ, ρ)，求某种准则意义下的最优随机决策函数 $\bar{\varphi}$。

根据以上陈述，在矿产预测的情况下，所谓统计决策问题就是，已知m种成矿远景类别和地质人员可能取的m种决策（即预测），n 维地质向量的集合X及其相对于m种远景类别的概率分布族 $\{P_i\}$，已知或未知m种远景类别的先验概率分布 $\bar{\mathbf{q}}$，求在上述条件下指导地质人员选择合理决策的符合某种准则的最优随机决策函数 $\bar{\varphi}$。

2. 最优性准则及统计决策问题的解法

选择随机决策函数的最优性准则通常有两种：贝叶斯准则和极小极大准则。

贝叶斯准则：在先验概率 $\bar{\mathbf{q}}\in Q$ 已知的条件下，选择使风险达到极小的 $\varphi\in\Phi$ 做为最优决策函数，即

$$\rho(\bar{\mathbf{q}}, \bar{\varphi})=\min_{\varphi\in\Phi}\rho(\bar{\mathbf{q}}, \varphi) \tag{2}$$

其中 $\bar{\varphi}$ 称为对于 $\bar{\mathbf{q}}$ 的贝叶斯策略。

极小极大准则：在先验概率未知时，选择对应于最不利的先验概率的贝叶斯策略，即在最不利的自然情况下能使风险达到极小的策略做为最优策略，即

$$\rho(\mathbf{q}^\circ, \varphi^\circ)=\min_{\varphi\in\Phi}\max_{q\in Q}\rho(\mathbf{q}, \varphi) \tag{3}$$

其中 $\mathbf{q}^\circ$ 为最不利的先验概率，φ° 为极小极大策略。

与上述两种准则相对应，统计决策问题有两种解法：贝叶斯解和极小极大解。以下我们分别对于 $m>2$ 和 $m=2$ 的两种情况给出这两种解法。

3. 大自然具有两种以上状态时统计决策问题的解法

先验概率 $\mathbf{q}\in Q$ 是一个m维向量

$$\mathbf{q}=(q_1, q_2, \cdots, q_m), \quad q_i\geqslant 0, \quad \sum_{i=1}^{m}q_i=1 \tag{4}$$

随机决策函数 $\boldsymbol{\varphi}$ 具有 $m\times l$ 维向量的形式：

$$\boldsymbol{\varphi}=(\varphi_{11},\varphi_{12},\cdots,\varphi_{1m},\varphi_{21},\varphi_{22},\cdots,\varphi_{2m},\cdots,\varphi_{l1},\varphi_{l2},\cdots\varphi_{lm});$$

$$\varphi_{rj}\geqslant 0,\ \sum_{j=1}^{m}\varphi_{rj}=1,\ r=1,2,\cdots l \tag{5}$$

风险函数 $\rho(\mathbf{q},\varphi)$ 的矩阵形式如下：

$$\rho(\mathbf{q},\varphi)=\mathbf{q}\boldsymbol{\rho}\boldsymbol{\varphi}^{\mathrm{T}} \tag{6}$$

其中矩阵 $\boldsymbol{\rho}$ 表示在大自然各种状态下对于不同的 n 维向量 $\mathbf{x}$ 取不同决策的风险，其形式为：

$$\boldsymbol{\rho}=(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,\cdots\cdots,\boldsymbol{\rho}_l) \tag{7}$$

其中 $\boldsymbol{\rho}_r$ 为 $m\times m$ 阶矩阵

$$\boldsymbol{\rho}_r=\begin{pmatrix} p_{r1}l_{11}, & p_{r1}l_{12},\cdots, & p_{r1}l_{1m}\\ p_{r2}l_{21}, & p_{r2}l_{22},\cdots, & p_{r2}l_{2m}\\ \cdots & \cdots & \cdots\\ p_{rm}l_{m1}, & p_{rm}l_{m2},\cdots, & p_{rm}l_{mm}\end{pmatrix}$$

$$r=1,2,\cdots,l。$$

(1) 贝叶斯解（$m>2$）

我们直接引用关于贝叶斯策略的下列定理。

定理：使策略 φ 成为对于先验分布 $\mathbf{q}$ 的贝叶斯策略的充要条件是，对于所有满足下式的 $\mathbf{x}_r$ 和 j，概率 $\varphi_j(\mathbf{x}_r)$ 都等于零：

$$t_j(\mathbf{x}_r)>\min_{l}\{t_1(\mathbf{x}_r),t_2(\mathbf{x}_r),\cdots,t_m(\mathbf{x}_r)\} \tag{8}$$

式中

$$t_j(\mathbf{x}_r)=\sum_{i=1}^{m}\mathrm{L}(i,j)\mathrm{P}_i(\mathbf{x}_r)q_i;\ i=1,2,\cdots,m,j=1,2,\cdots,l。$$

设已知某个 $\bar{\mathbf{q}}\in\mathrm{Q}$，以 $\bar{\mathbf{q}}$ 乘 ρ 得

$$\mathbf{t}=(t_{11},t_{12},\cdots,t_{1m},t_{21},t_{22},\cdots,t_{2m},\cdots,t_{lm}) \tag{9}$$

由 (8) 式可知，(5) 式给出的贝叶斯策略 $\bar{\varphi}$ 只能有等于 0 和 1 的元素，且对每个给定的 r，只有与

$$\min\{t_{r1},t_{r2},\cdots,t_{rm}\},\ r=1,2,\cdots l \tag{10}$$

相对应的那个元素 $\bar{\varphi}_{rj}$ 能够等于 1。

因此，贝叶斯解给出的是一个整数组 $\{j_r^0\}$，$r=1,2,\cdots l$，其中每一个整数都是 (10) 式中的最小元素的号码，这是在观测每个组合 $\mathbf{x}_r$ 时采取的最优决策的号码。而贝叶斯风险(即取上述决策的风险)由下式确定：

$$\rho(\bar{\mathbf{q}},\bar{\varphi})=\sum_{r=1}^{l}t_{j_r^\circ r} \tag{11}$$

(2) 极小极大解（$m>2$）

通过解下列初始的和对偶的线性规划问题可以求得在大自然最不利的先验概率条件下统计人员的最优策略(极小极大策略) $\boldsymbol{\varphi}^*$ 及大自然的最不利先验概率 $\mathbf{q}^*$。

初始线性规划问题：

$$\left.\begin{aligned} &min\ \mathbf{CU} \\ &s.t.\ \mathbf{TU} \geqslant \mathbf{B} \\ &\quad\ \mathbf{U} \geqslant 0 \end{aligned}\right\} \tag{12}$$

对偶线性规划问题：

$$\left.\begin{aligned} &max\ \mathbf{VB} \\ &s.t.\ \mathbf{VT} \leqslant \mathbf{C} \\ &\quad\ \mathbf{V} \geqslant 0 \end{aligned}\right\} \tag{13}$$

式中 $\mathbf{U} = (\boldsymbol{\varphi}, p)^{\mathrm{T}}$ 为 $m \times l + 1$ 维列向量；其中 $\boldsymbol{\varphi}$ 为随机决策函数，$m \times l$ 维行向量，p 为标量；

$\mathbf{V} = (\mathbf{q}, \mathbf{t})$ 为 $m + l$ 维行向量；其中 $\mathbf{q}$ 为大自然的先验概率，m 维行向量，$\mathbf{t}$ 为 l 维行向量；

$$\mathbf{B} = (\underbrace{0, \cdots, 0}_{m\text{个}}, \underbrace{1, \cdots, 1}_{l\text{个}})^T \text{——} m + l \text{ 维列向量；}$$

$$\mathbf{C} = (\underbrace{0, \cdots 0}_{m\times l\text{个}}, 1) \text{——} m \times l + 1 \text{ 维行向量；}$$

$$\mathbf{T}\text{——}(m + l) \times (m \times l + 1) \text{ 阶约束矩阵}$$

$$\mathbf{T} = \begin{pmatrix} -\boldsymbol{\rho} & \mathbf{I}_m \\ \mathbf{S} & 0 \end{pmatrix}$$

其中 $\boldsymbol{\rho}$——$m \times ml$ 阶风险矩阵，如 (7) 式；

$\mathbf{S}$——$l \times ml$ 阶矩阵

$$\mathbf{S} = \begin{pmatrix} \mathbf{I}_m^{\mathrm{T}} & 0 \cdots 0 \\ 0 & \mathbf{I}_m^{\mathrm{T}} \cdots 0 \\ \cdots & \cdots\cdots \\ 0 & \cdots\cdots\ \mathbf{I}_m^{\mathrm{T}} \end{pmatrix}$$

$\mathbf{I}_m$——由 1 组成的 m 维列向量。

用单纯形法[5,6]可以求得上述初始的和对偶的线性规划问题 (12) 和 (13) 的最优解 $\mathbf{U}^*$ 和 $\mathbf{V}^*$。由对偶线性规划的性质[4]可知：

$$\mathbf{CU}^* = \mathbf{V}^*\mathbf{B}$$

$$\mathbf{CU}^* = (0, \cdots, 0, 1)\begin{pmatrix} \boldsymbol{\varphi}^{*\mathrm{T}} \\ p^* \end{pmatrix} = p^*$$

$$\mathbf{V}^*\mathbf{B} = (\mathbf{q}^*, \mathbf{t}^*)\begin{pmatrix} 0 \\ \mathbf{I}_l \end{pmatrix} = \sum_{i=1}^{l} t_i^*$$

故

$$p^* = \sum_{i=1}^{l} t_i^* = v.$$

将 $\mathbf{U}^*$ 和 $\mathbf{V}^*$ 分别代入 (12) 及 (13) 式，并加以展开可得到

$$\left.\begin{aligned} &\mathbf{q}\boldsymbol{\rho}\boldsymbol{\varphi}^{*\mathrm{T}} \leqslant v \leqslant \mathbf{q}^*\boldsymbol{\rho}\boldsymbol{\varphi}^{\mathrm{T}} \\ &v = \mathbf{q}^*\boldsymbol{\rho}\boldsymbol{\varphi}^{*\mathrm{T}} \end{aligned}\right\} \tag{14}$$

这正是以 $(\mathbf{q}^*, \boldsymbol{\varphi}^*)$ 为解的矩阵博奕的基本性质。$\boldsymbol{\varphi}^*$ 是极小极大策略，$\mathbf{q}^*$ 是大自然

最不利的先验概率，而 v 是博奕值。

4. 大自然只有两种状态时统计决策问题的解法

在这种情况下，统计决策问题的解法可以简化。对于矿产预测问题可以约定，损失 l_{ij} 满足下列条件：

$$l_{ij} > l_{ii} \text{ 及 } l_{ij} > l_{jj},\ i \neq j。$$

(1) 贝叶斯解 (m = 2)

对于满足下式

$$t_r = P_{2r}/P_{1r} \leqslant w_1 q/(1-q) = C_q \tag{15}$$

的脚标 r(n 维向量 $\mathbf{x}_r$ 的号码)取决策 1，其余情况取决策 2。上式中 t_r 为风险函数值；P_{ir} 为对应于向量 $\mathbf{x}_r$ 概率分布 P_i 的值，即

$$P_{ir} = P_i(\mathbf{x}_r),\ r = 1, 2, \cdots, l,\ l = k^n;$$

w_1 为由损失函数决定的系数

$$w_1 = (l_{12} - l_{11})/(l_{21} - l_{22})$$

在标准化损失函数

$$L = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

时，$w_1 = 1$；q 系大自然为第一种状态时的先验概率。C_q 称为贝叶斯决策规则。

贝叶斯策略 ψ 相对于给定的先验概率 $\mathbf{q}$ 使风险达到极小，后者称为贝叶斯风险，即

$$\rho(\mathbf{q}, \psi) = q[l_{11} + (l_{12} - l_{11})\alpha_q] + (1-q)[l_{22} + (l_{21} - l_{22})\beta_q] \tag{16}$$

在标准化损失函数情况下，贝叶斯风险为

$$\rho(\mathbf{q}, \psi) = q\alpha_q + (1-q)\beta_q \tag{17}$$

式中

$$\left.\begin{aligned} \alpha_q &= 1 - \sum_{r \in R} P_{1r} = \sum_{r \notin R} P_{1r} \\ \beta_q &= \sum_{r \in R} P_{2r} = 1 - \sum_{r \notin R} P_{2r} \end{aligned}\right\} \tag{18}$$

其中脚标集 R 由满足 $t_r \leqslant C_q$ 的脚标 r 组成。

(2) 极小极大解 (m = 2)

如果将向量 $\mathbf{x}_r$ 按下列方式排列

$$t_1 \leqslant t_2 \cdots\cdots \leqslant t_l,$$

则极小极大策略就在于选择一个满足不等式

$$\rho_{1s_0} \geqslant \rho_{2s_0} \text{ 和 } \rho_{1s_0+1} < \rho_{2s_0+1} \tag{19}$$

的号码 s_0。对于 $r < s_0$，取决策 1；对于 $r > s_0$，取决策 2；而 $r = s_0$ 时，以概率 r 取决策 2。这里，ρ_{is}——大自然为第 i 种状态时，以 s 为界决策的风险，即 $\rho_{is} = \rho(i, \psi^s)$，它由下式决定：

$$\left.\begin{aligned} \rho(1, \psi^s) &= l_{11} + (l_{12} - l_{11})\alpha^s \\ \rho(2, \psi^s) &= l_{22} + (l_{21} - l_{22})\beta^s \end{aligned}\right\} \tag{20}$$

概率 γ 由下式确定：

$$\gamma = \frac{w_2 + w_1\alpha^{s_0+1} - \beta^{s_0+1}}{(w_1\alpha^{s_0+1} - \beta^{s_0+1}) - (w_1\alpha^{s_0} - \beta^{s_0})} \tag{21}$$

式中

$$\alpha^s = 1 - \sum_{r=1}^{s} P_{1r}, \quad \beta^s = \sum_{r=1}^{s} P_{2r};$$

$$w_2 = (l_{11} - l_{22})/(l_{21} - l_{22})。$$

最不利的先验概率 $\mathbf{q}^\circ$ 通过解下列方程求得：

$$w_1 q^\circ/(1 - q^\circ) = t_{s_0+1} \tag{22}$$

博奕值 v 按下式计算

$$v = \gamma \rho_{1s_0} + (1 - \gamma)\rho_{1s_0+1} \tag{23}$$

5. 概率分布族 $\{P_i\}$ 的逼近方法

做为统计决策问题的已知条件之一的概率分布族 $\{P_i\}$ 在实践中往往是未知的，只能通过抽样估计的方法加以确定。这就是说，要给定 m 类标准个体集 $A_1, A_2, \cdots A_m$ 的描述 $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(m)}$，它们是集 X 的 m 个已知子集。

由于涉及到条件概率，联合分布 $P_i(\mathbf{x}) = P_i(x_1, x_2, \cdots, x_n)$ 在实践中是很难估计的。因此，必须通过附加某种限制来实现对 $P_i(\mathbf{x})$ 的逼近。最常用的 $P_i(\mathbf{x})$ 的逼近方法是一阶逼近。这种方法是以诸变量的统计独立性为前提的。在此情况下，

$$P_i(\mathbf{x}) = \prod_{j=1}^{n} P_i(x_j) \tag{24}$$

式中

$$P_i(x_j) = \begin{cases} P_{ij1}, & 当\ x_j = v_1 \\ P_{ij2}, & 当\ x_j = v_2 \\ \vdots & \\ P_{ijk}, & 当\ x_j = v_k \end{cases}$$

$$\sum_{s=1}^{k} P_{ijs} = 1$$

其中

$$P_{ijs} = P_i(x_j = v_s), \quad s = 1, 2, \cdots, k$$

v_s——x_j 可能取的第 s 个离散值。

数组 $\{P_{ijs}\}$ 的子样估计通常表示成 m 个 $n \times k$ 阶矩阵 $\mathbf{R}_i = \{r_{ijs}\}$，其中 r_{ijs}——第 j 个变量在第 i 个子样中取第 s 个值的频率，$s = 1, 2, \cdots k$；$j = 1, 2, \cdots n$；$i = 1, 2, \cdots, m$。对于所有的 i, j 有 $\sum_{s=1}^{k} r_{ijs} = 1$。

当 $k = 2$ 时，$P_i(x_j)$ 可以按下式计算：

$$P_i(x_j) = (p_{ij})^{x_j} \cdot (q_{ij})^{1-x_j} \tag{25}$$

式中

$$p_{ij} = \hat{P}_i(x_j = 1),\ q_{ij} = \hat{P}_i(x_j = 0)\ p_{ij} + q_{ij} = 1。$$

数组 $\{p_{ij}\}$ 和 $\{q_{ij}\}$ 的子样估计可以表示成 m 个 $n \times 2$ 阶矩阵 $\mathbf{R}_i\{r_{ijs}\}$，其中 r_{ijs} 同前，但 $s = 0, 1$。

6. 先验概率 q 的估计方法

在矿产预测的条件下，大自然的先验概率 $\mathbf{q}$ 是指矿产的成矿概率，或矿点在空间的概率分布。在一般情况下，成矿概率是未知的，通常可采用下列两种方法之一进行估计。

(1) 理论分布配赋法

在工作区有足够数量已知矿床、矿点的情况下，可根据矿点概率分布类型的假设，对矿点(或储量)按面积的经验分布资料进行理论配赋，可求得不同矿点数(或储量)区间的产出概率，以此做为先验概率 $\mathbf{q}$。

目前已提出的矿点的空间概率分布类型有以下几种：波松分布、负二项分布、涅曼-斯科特分布、齐波夫分布等。笔者根据以往的经验认为，有色及稀有金属矿产似应服从负二项分布。

(2) 主观概率反馈修正法

在工作区缺乏足够数量矿床矿点的情况下可由找矿经验丰富、对工作区地质条件熟悉的地质人员，根据同类型地区找矿的经验，对成矿概率进行主观估计。然后以此先验概率用贝叶斯法对工作区的各类已知单元进行决策分类，根据分类结果与实际类别的对比，对上述先验概率反复加以修正，使之对已知单元的分类达到最佳效果。此时之先验概率即为所求的先验概率。

限于篇幅的原因，本文不拟讨论先验概率的具体算法。

二、原始数据的预处理

1. 主成分分析及变量信息量的估计

为了保证变量的独立性，需用主成分分析方法[7]预先将原始变量转换成互相独立的主因子，做为统计决策分析的起始变量。

为了对变量进行排队筛选和连续性数据离散化，这里介绍两种对变量信息量进行估计的方法。

(1) 库尔巴克法

在 $m = 2$ 的情况下，信息量按库尔巴克公式直接计算：

$$I = \sum_{s=1}^{k} (m_s/N_1 - n_s/N_2) l_n [N_2 m_s/(N_1 n_s)] \tag{26}$$

式中

N_1，N_2——第一、二子样的容量；

m_s，n_s——第一，二子样中变量取第 s 个离散值的个体数；

k——变量取值的个数，对于二态变量 $k = 2$。

当 $m > 2$ 时，采用下述方法之一计算信息量：

a. 将一个变量转化为m个二态变量,分别按(26)式计算信息量;

b. 先按(26)式计算每对子样的信息量,而后求其平均值做为变量之平均信息量。

(2) 值组对应法

实践表明,库尔巴克法存在三点不足之处,即:a. (26)式中之m_s和n_s不能有零值存在,否则无法计算信息量;b.库尔巴克信息量没有确定的上限,以致在筛选变量时无法给出一个信息量的阈值;同时,也无法在不同批数据之间比较变量的信息量;c.库尔巴克公式只适用于$m=2$的情况,在$m>2$时,该式无法直接应用。

鉴于上述问题,作者提出下列信息量公式:

$$\mathrm{I}=\frac{1}{m}\sum_{i=1}^{m} n_i^i/\mathrm{N}_i \tag{27}$$

式中

N_i——第i个子样的容量;

n_i^i——第i个子样中变量取相应的第i个值的个体数;

m——既是子样数,又是变量取值个数。

上述公式要求$k=m$,即变量取值个数与分类的组数相等,而且值与组的顺序相对应。该公式克服了上述三个缺点,它对参加计算的诸量没有附加限制,其值变化于(0, 1)区间,且适合于$m\geqslant 2$的情况。

2. 连续型数据的最佳离散化

连续型数据的离散化是应用离散型统计决策分析时必不可少的预备步骤。这里介绍的最大信息量法能最大限度利用有用的分类信息。

$m=2$的情况。设$\mathbf{x}_j^{(i)}=(x_1^{(i)}, x_2^{(i)}, \cdots, x_{N_i}^{(i)})$为第$i$个子样中变量$x_j$的连续值向量,$i=1, 2$。要求找到一个离散化的临界值$c$,使信息量达到极大。

为使用(25)或(26)式,临界值c应满足不等式

$$a=\max_i \min_r \mathbf{x}_r^{(i)}\leqslant c\leqslant \min_i \max_r \mathbf{x}_r^{(i)}=b$$

若$a>b$,则不可能有满足上式的c。这时$\mathbf{x}_j^{(1)}$与$\mathbf{x}_j^{(2)}$不相交,可任选$c\in[a, b]$作为临界值,例如,$c=(a+b)/2$。

若$a<b$,选定(根据经验)步长$\Delta>0$,将区间$[a,b]$分割为t个长度为Δ的子区间,其中$t=(b-a)/\Delta$。对所得$t-1$个分界点计算信息量。根据下式确定的分界点c:

$$\mathrm{I}(c)=\max\{\mathrm{I}(c_1), \mathrm{I}(c_2), \cdots, \mathrm{I}(c_{t-1})\}$$

即所求最佳离散化临界值。

当$m>2$时可采用上述$m=2$时的方法分别确定各离散临界值。

作者根据上述数学模型设计了相应算法和 FORTRAN77 计算程序,经过实用,效果良好。

三、应用实例

预测区位于一复背斜南翼。轴部呈北西295°展布,核部为晚太古代花岗杂岩体所占

据。并保留有部分鞍山群残留体。该岩体经受了多次后期交代作用，它以红色似斑状花岗岩为主体，边部有白色混合花岗岩。复背斜两翼以下元古界辽河群为主。辽河群与下伏花岗杂岩体呈构造和混合交代接触。区内构造发育，具多期性和继承性特点。主要构造线呈北西西向展布，地层与岩体接触带为一条规模较大的挤压焊接破碎带。复背斜两翼地层曾遭受多次褶皱变形。区内铀矿化发育，已发现有铀矿床产出。矿化受白色混合花岗岩、浪子山组底部层位、断裂破碎带、回曲构造及碱交代作用控制。

预测区面积约 150 平方公里。预测单元沿实测剖面线划分，单元大小沿走向 500 米，倾向 250 米。全区划分为 440 个基本单元，其中包括有矿单元 13 个，无矿单元 15 个。

参与预测的原始资料系本区地质测量及物化探。水化学测量资料。初选变量 23 个，其库尔巴克信息量及离散化界限列于表 1。按信息量排队后，选出信息量大于 1.5 的变量8个(表中带 * 者)参加预测计算。预测方法采用离散型统计决策分析。

首先用主成分分析将 8 个原始变量转换成 6 个相互独立的主因子，累积因子贡献达 90.93%。主因子矩阵如表 2 所示。

表 1　各类变量信息量及离散化界限表

Table 1　Information contents and levels for discretization of variables

序号	变量类别	变量名称		信息量	离散化界限(取值为 1)	按信息量排队
1	地质变量		下元古界底部第一岩段 Ll_1 出露面积	0.59	> 0(m²)	15
2			Ll_1 中石英岩颜色	1.32	灰或灰白	9
3			下元古界底部第一、二岩段中石英岩和片岩互层*	1.96	有	4
4			底部层厚度*	1.50	1—700(m)	8
5			单元至白色混合花岗岩距离*	>1.51	< 900(m)	6
6			单元至闪长岩脉距离	1.23	> 2000(m)	10
7			单元至变辉长岩脉距离	0.77	< 650(m)	13
8			单元附近地层产状有无变化*	1.57	有	5
9			单元附近有无回曲构造或脉岩密集*	2.00	有	3
10			单元内地层是否属二级变质程度	1.23	是	10
11	水化学变量		单元周围水中铀偏高点个数*	2.16	> 3	2
12			单元周围水中铀偏高点及异常点个数	1.13	> 4	11
13	放射性物化探变量	能谱测量	单元铀含量趋势平均值	0.49	>0.68(In)	17
14			单元铀含量正残差平均值	0.32	>0.55(In)	18
15			单元钍、铀比值趋势平均值*	2.70	1.87—2.20(In)	1
16			单元钍、铀比值负残差平均值	0.52	<−0.35(In)	16
17			单元铀含量组合特征值	0.83	>26	12
18			单元钍含量组合特征值	0.69	>38	14
19			单元钍、铀比值变异系数	0.83	>47%	12
20		径迹测量	单元内高场以上面积*	1.51	>1300(m²)	7
21			单元内异常场面积	1.23	>200(m²)	10
22		伽玛测量	单元内偏高场以上面积	0.17	>21500(m²)	19
23			单元组合特征值	0.17	>48	19

通过主因子估计求得第一、二子样各单元的主因子得分，并以最大信息量法进行离散

表 2 主因子载荷表

Table 2 Loadings of variables on principal components

变量编号	主因子编号					
	1	2	3	4	5	6
1	0.573	−0.100	0.759	−0.011	−0.208	0.075
2	0.685	−0.312	−0.388	−0.279	−0.272	−0.054
3	0.640	0.570	−0.058	−0.068	−0.279	−0.131
4	0.697	−0.441	0.179	−0.302	0.262	−0.170
5	0.840	−0.191	−0.124	0.190	−0.054	0.253
6	0.832	0.177	0.020	0.376	0.114	−0.253
7	0.906	−0.186	−0.191	0.167	0.111	0.106
8	0.613	0.644	0.006	0.282	0.247	0.160
因子贡献	4.287	1.143	0.814	0.458	0.353	0.219
累积贡献(%)	53.59	67.87	78.05	83.78	88.19	90.93

化，将各主因子变换成二态变量。离散化结果列于表 3。各主因子之信息量及离散化临界值示于表 4 左侧。

为了计算两个子样的频率分布，首先统计两个子样 0 与 1 的频率，组成矩阵 $\mathbf{R}_1$ 和 $\mathbf{R}_2$。结果列于表 4 右侧。然后按 (23)，(24) 式计算频率分布 P_1 和 P_2。

表 3 主因子得分离散化结果表

Table 3 Results of discretizating scores on principal components

子样号	单元号	主因子编号						子样号	单元号	主因子编号					
		1	2	3	4	5	6			1	2	3	4	5	6
第一子样	1	1	1	0	0	1	1	第二子样	14	0	1	0	1	1	0
	2	1	1	1	1	0	0		15	0	1	0	0	1	0
	3	1	1	0	1	1	0		16	0	1	0	0	1	0
	4	1	1	0	1	0	0		17	0	1	0	1	0	0
	5	1	1	1	1	0	0		18	0	0	0	1	1	0
	6	1	1	1	1	0	0		19	0	1	0	0	1	1
	7	1	1	1	1	0	0		20	0	0	0	0	0	0
	8	1	1	1	1	0	0		21	1	0	0	1	1	1
	9	1	1	1	1	1	0		22	1	1	0	0	1	1
	10	1	1	1	1	1	0		23	0	1	0	1	0	1
	11	1	0	1	1	0	0		24	0	0	1	1	0	0
	12	1	1	1	0	0	0		25	0	1	1	0	1	1
	13	0	1	0	0	1	0		26	0	0	1	0	1	0
									27	0	0	1	0	1	0
									28	0	0	0	1	1	0

损失函数采用标准化形式，即

$$L = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

故 $w_1 = 1$，$w_2 = 0$。

表 4　诸因子信息量、离散化临界值、矩阵 R_1 和 R_2 表

Table 4　Information contents and levels for discretization of principal components, matrixes R_1 and R_2

因子编号	离散化临界值	信息量	按信息量排队	R_1		R_2	
				"0" 的频率	"1" 的频率	"0" 的频率	"1" 的频率
1	−0.099	3.44	1	0.08	0.92	0.87	0.13
2	−0.497	0.92	2	0.08	0.92	0.47	0.53
3	0.021	0.78	3	0.31	0.69	0.73	0.27
4	−0.233	0.40	6	0.23	0.77	0.53	0.47
5	−0.325	0.52	4	0.62	0.38	0.27	0.73
6	0.672	0.46	5	0.92	0.08	0.67	0.33

然后，按 (15) 式的对数形式

$$t_r = \ln P_{2r} - \ln P_{1r} \leqslant \ln[q/(1-q)] = C_q$$

计算 $l = 2^6 = 64$ 种变量值组合 $\mathbf{x}_r$ 的风险函数值 t_r，$r = 1, 2, \cdots 64$，以及贝叶斯决策规则 C_q。当给定 $q = 0.5$ 时，$C_q = 0$；$q = 0.1$ 时，$C_q = -2.197$。根据每个单元的主因子二态值组合，可查得其 t_r 值。然后与 C_q 比较，若 $t_r \leqslant C_q$，单元被认为有成矿远景，否则无成矿远景。表 5 列出了已知单元的 t_r 值及 $q = 0.5$ 时的贝叶斯决策结果。

表 5　已知单元的贝叶斯决策结果表

Table 5　Results of Bayes'decision of know units

第一子样(有矿组)			第二子样(无矿组)		
单元号	t_r 值	远景类	单元号	t_r 值	远景类
1	1.334	II	14	2.562	II
2	−5.099	I	15	3.899	II
3	−1.795	I	16	3.899	II
4	−3.276	I	17	1.080	II
5	−5.099	I	18	4.913	II
6	−5.099	I	19	5.691	II
7	−5.099	I	20	4.769	II
8	−5.099	I	21	2.348	II
9	−3.617	I	22	1.334	II
10	−3.617	I	23	2.872	II
11	−2.748	I	24	1.609	II
12	−3.761	I	25	3.869	II
13	3.899	II	26	4.428	II
			27	4.428	II
			28	4.913	II

已知单元决策结果与其原属类别对比如表 6。错判两个单元，其中 13 号单元矿化情况不好，不应选为已知有矿单元。

待预测单元 412 个，预测结果 I 类单元 32 个，II 类单元380个。

贝叶斯风险为 0.087，即理论错判率为 8.07%。实际错判率为 7.7%、二者吻合较好。

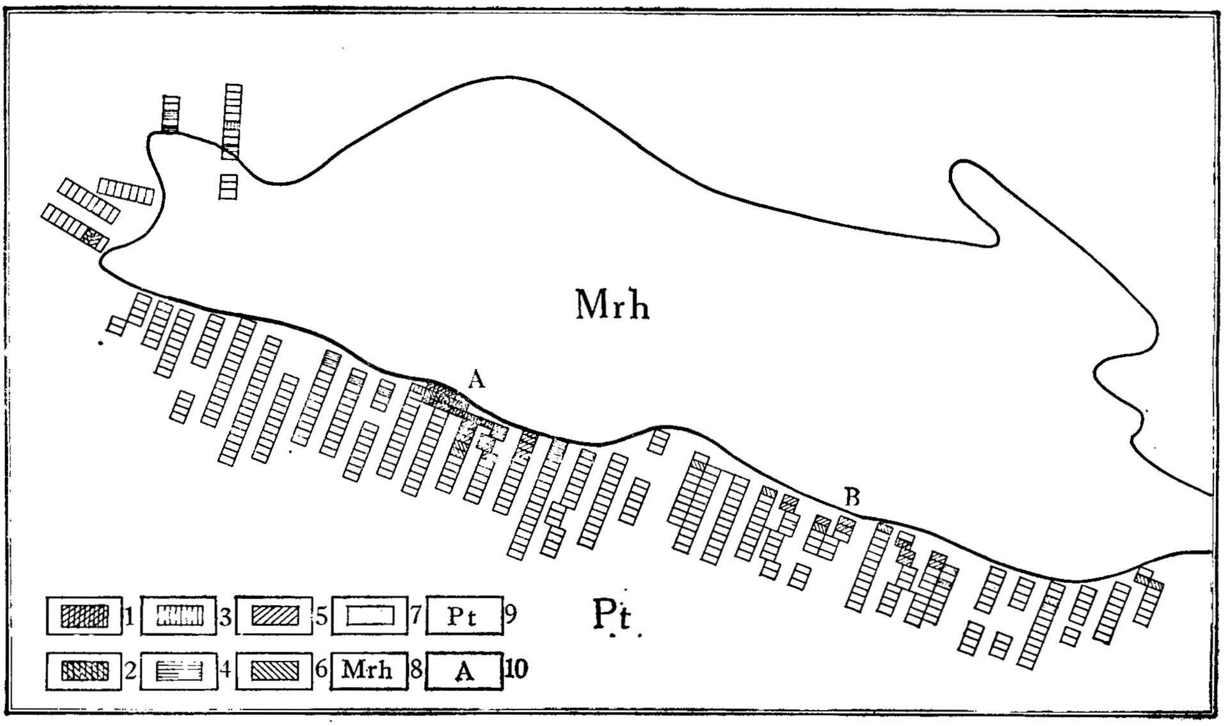
表 6

Table 6

原属类别 / 决策类别	I	II
I	11	0
II	2	15

极小极大解的情况如下。64 个向量 $\mathbf{x}_r$ 按 t_r 值的递增顺序排列后，求得 $s_0 = 20$，相应的 t_r 值为 -0.147，t_r 小于此值的单元划为 I 类。$s_0 + 1 = 21$，相应的 t_r 值为 -0.003，t_r 大于此值的单元划为 II 类，t_r 值等于 -0.003 的单元以 $\gamma=0.85$ 的概率划为 II 类。博奕值(即理论风险) $v = 0.0807$。大自然最不利的先验概率对于含矿类 $q^0 = 0.499$，无矿类 $1 - q^0 = 0.501$。预测结果与 $q = 0.5$ 时的贝叶斯解相同。

为了对预测所得 I 类（有远景）单元进一步分级，按 $q = 0.1$，$C_q = -2.917$ 重新进行了贝叶斯决策。所得 I 类单元做为一级远景单元。原按 $q = 0.5$ 预测之 I 类单元除一级单元外定为二级远景单元。据此，本区计得一级远景单元 32 个(包括已知有矿单元 10 个)，二级远景单元 11 个(包括已知有矿单元 1 个)，无远景单元 397 个。如图所示。



三〇七五地区铀矿远景预测图

Prediction map of mineralizing prospect of uranium of 3075 region

1.已知有矿单元判为一级远景单元；2.已知有矿单元判为二级远景单元；3.已知有矿单元判为远景单元；4.已知无矿单元判为无远景单元；5.一级远景单元；6.二级远景单元；7.无远景单元；8.太古代红色混合花岗岩；9.早元古代辽河群；10.成矿远景片 A(B 亦然)

根据以上决策结果，本区划分出 A，B 两个较好的成矿远景片。A片是三〇七五矿床外围（该矿床向东向南延伸部分），控制面积约 4 平方公里已发现铀矿床一处，矿点两处，地质、物化探及水化学条件均为有利地段。B片是二〇六矿点外围，控制面积约 3 平方公里，已发现铀矿点一处，地质、物化探及水化条件均为有利地段。

为了对比方法之效果，作者对本区原始数据应用两类逐步判别分析进行了预测。结果得 I 类单元 52 个，II 类单元 388 个，原属 I 类的 13 号单元也判为 II 类。统计决策与逐步判别的预测结果对比情况见表 7。二者相符者占单元总数的 95.7%，不符者占 4.3%。

表 7　两种方法对比表

Table 7　Contrast between both methods

逐步判别类别 \ 统计决策类别	I	II
I	38	14
II	5	383

由以上与已知类别及判别分析对比结果可见，统计决策方法完全适用于矿产预测工作。此外，它还具有下列独特的优点：

（1）损失函数的参与为使预测工作与经济效益联系起来创造了条件；

（2）理论风险值提供了对预测结果的可靠性估计，并可提供不同可靠程度的预测结果。

参　考　文　献

[1] Бугаец, А. Н., Дуденко, Л. Н., 1976, Метематические методы при прогнозировании месторождений полезных ископаемых. Недра, Лениигрэд, стр. 32—69, 154—159.

[2] Дуденко, Л. Н.. 1973, Минимаксные стратегии в статистических играх, соответствующих задачам распознавания. Тех. кибернетика, №6, стр. 61—68.

[3] 瓦尔特，A. 1960，统计决策函数。王福保译，第 1—30 页。上海科学技术出版社。

[4] 中国科学院数学研究所第二室，1960，对策论（博奕论）讲义。第 13—68，143—155 页，人民教育出版社。

[5] Гольштейн, Е. Г., Юдин, Д. Б., 1966, Линейное программирование и теория игр, в кн. Новые направления в линейном программировании. стр. 83—123. Изд. Советское радио, Москва.

[6] 鲁恩伯杰，D. G. 1980，线性与非线性规划引论。夏尊铨译，第 30—100 页。科学出版社。

[7] 王学仁，1980，地质数据的多变量统计分析。第 249—305 页。科学出版社。

THE STATISTICAL DECISION METHOD FOR PREDICTION OF MINERAL RESOURCES

Yin Di and He Guoxin

Abstract

Prediction of minera lresources is a statistical decision problem. Classes of mineralization prospect are regarded as conditions existed in the nature and the predictions of mineralization prospect made by a geologist may be considered as decisions of a decider. In this case the statistical decision problem is stated as follows: given m the classes of mineralization prospect and m, decision of geologists, set X of all possible n-dimensional geological vectors and class {Pi} of their probability distributions for m prospect classes, known or unknown prior probability **q** of m prospect classes, to find an optional decision function φ corresponding to some principle, which can guide geologist to select optional decision. In correspondence with two optimality principles are given the Bayes' and minimax solution of statistical deoision problem for the cases of $m>2$ and $m=2$ respectively.

In the paper the technigue of the first order approximation to class {Pi} of probbility distributions and methods of pre-proccesing for original data are described.

Application of the statistical decision method in the prediction of uranium mineral resources for the 3075 region, for example, is very successful.